Application of Neural Networks to Large-Scale Cloud Pattern Recognition

J. E. Peak
Computer Sciences Corporation
Monterey, CA 93943-5006

Prepared for
Atmospheric Directorate
Monterey, CA 93943-5006

ABSTRACT

A Multi-layer Perceptron Neural Network methodology is used to classify eight types of large-scale cloud patterns. The data are taken from GOES-W visible images from Oct. 1 - Dec. 31, 1983. Large-scale features are previously identified by a human expert to provide a data set for supervised learning. Discriminant Analysis is used to reduce the set of network inputs and as a comparison classification methodology. In three different tests, the neural network technique classifies the cases with consistently higher accuracy than Discriminant Analysis. The problem of image segmentation is addressed in a preliminary test of the Hierarchical Stepwise Optimization algorithm.

ACKNOWLEDGMENTS

The author gratefully acknowledges the support of the sponsor, Office of Naval Technology, Code 22, Mr. James Cauffman, Program Element 62435N, for making this effort possible.

TABLE OF CONTENTS

APPLICATION OF NEURAL NETWORKS TO LARGE-SCALE CLOUD PATTERN RECOGNITION

1. Introduction

In a previous study Peak (1990) proposed the use of neural networks for the interpretation of certain cloud features on satellite images. That paper includes a preliminary experiment in which large-scale cloud patterns (fronts, cirrus and vortices) on GOES infrared images are distinguished using a neural network. The preliminary experiment was designed to explore the use of neural nets with simple areal cloudiness percentages as inputs. The success of that simplified approach led to the proposal of a more advanced approach using cloud-type inputs instead of cloud percentages.

The purpose of this paper is to document this new approach for neural classification of large-scale cloud features. The reader is referred to Peak (1990) for background information on neural networks including a comparison of several types of neural networks and a mathematical description of the multi-layer perceptron nets used here.

The data used in this study are described in the next section. An initial screening of the network inputs using stepwise discriminant analysis is described in Section 3. In Section 4 the neural net derivation will be described including the network results on test data. In Section 5 the problem of automated image segmentation is addressed. Finally, the conclusions of this study and suggestions for future research will be presented.

2. Data description

As described in Peak (1990), multi-layer perceptron neural networks require a set of training cases with known outputs.

These cases are used in a supervised-learning mode to derive the network weights. In this section the data used for training and testing the neural net will be presented.

As in the previous study (Peak, 1990), GOES imagery is used because of its wide field of view. Since TESS* does not receive GOES imagery, the problem of using polar-orbiting imagery must eventually be addressed. However, for these preliminary studies the more important issue is to determine the feasibility of using neural nets to accomplish image feature classifications. Therefore it was decided, with the approval of the User Project Manager, to continue using GOES data for these initial studies.

In this study archived GOES-West images were acquired from the period October-December 1983. The 2045 UTC visible and IR images were selected every three days beginning Oct. 1, yielding 31 western North Pacific scenes containing various large-scale cloud features. Because the multi-layer perceptron neural network is trained using supervised learning (Peak, 1990), it was necessary that the large-scale features in these images be classified a priori. In addition, the types of clouds present had to be determined to provide inputs for the network. The ideal method of determining the cloud types would be to use an objective cloud classification scheme. Unfortunately, the methods currently under development have not yet reached a sufficient level of capability to be used for this experiment. Therefore it was decided, again with the approval of the User Project Manager, to use cloud and feature classifications performed by an image interpretation expert. In future studies, these steps would have

* Tactical Environmental Support System

to be accomplished by automated processes. However, for the purpose of this study this approach can be likened to a "perfect prog" because an automated approach would probably include some erroroneous classifications.

The large-scale feature identification and cloud-typing of these images was kindly performed by Mr. R. Fett of NOARL-W. The eight large-scale feature types he identified and the number of occurrences of each are presented in Table 1. Some features that were labeled differently on different images have been combined into the same category. For example, features labeled "Frontal band" are considered the same type as those labeled "Cold Front." Similarly, "Trough" and "Upper cold low" were combined, as were "Stratocumulus" and "Open cells," "Tropical cyclone" and "Hurricane," and "Cirrus" and "Jet cirrus." Notice also that there is a distinction between frontal bands with a vortex at the northern end and those with no vortex (Table 1). There were also a few other features that were excluded because they appeared only once in the data set.

Table 1. Large-scale cloud feature types and number of occurrences of each type identified in the GOES-W image set.

FEATURE	NUMBER
Frontal band/Cold front (no vortex)	37
Frontal band/Cold front (with vortex)	10
Trough/Upper cold low	12
Stratocumulus/Open cells	53
Fog	9
Tropical cyclone/Hurricane	8
Cirrus/Jet cirrus	7
ITCZ	36
Total	172

The next task is to define the inputs to the neural net. As in the first experiment (Peak, 1990) the procedure is to keep the data used as simple as possible. If the results indicate that the information in the network inputs is inadequate, then more complex predictor information can always be included later.

A specific goal of this study is to use the type of cloud as an input for the network. The cloud type categories identified by Mr. Fett are "High," "Low," "Multi-layer" and "Stratocumulus." Obviously, the "Stratocumulus" feature-type (Table 1) contains that particular cloud-type. However, it would be meaningless to identify the "Stratocumulus" cloud feature by telling the neural network that it is made up of stratocumulus clouds. Since "Stratocumulus" is the only meteorological cloud type included, it was decided that the cloud type predictor should be "Low" for that feature.

Each cloud type is assigned to a corresponding network input. If the cloud type is present in the feature at hand, the input is assigned the value 1.0; otherwise the input is 0.0. If a feature contains regions of different types of clouds, more than one input could be assigned the 1.0 value. For example, some frontal bands have multi-layer clouds at their northern end and low clouds at their southern end.

It seems reasonable that the identification of a cloud feature requires some information about its shape. As a very rough, first estimate of shape it was decided to include the zonal and meridional dimensions (in degrees longitude and latitude, respectively) of each feature. Intuitively, this shape measure should

probably be inadequate for some features, but as will be seen later, it suffices quite well for this experiment when combined with the other information at hand.

Because it was found to be important in the preliminary study (Peak, 1990), the northmost latitude of the cloud feature is also included as an input. When combined with the zonal and meridional dimensions and the three cloud types, the northmost latitude provides a total of six inputs for the neural network. There are many other potential inputs that could be included, but these six were considered a good set with which to start the analysis. The complete set of 172 cases including input values and feature type is presented in Table 2.

Table 2. Initial case set for large-scale neural net development. Date and Label are included for reference, Del-x and Del-y denote zonal and meridional dimensions (degrees/100), Mult, High and Low denote cloud type presence, North is northmost extent of the feature (degrees latitude/100) and Type is large-scale feature class.

Date	Label	Del-x	Del-y	Mult	High	Low	North	Type
Oct 1	a-1	0.55	0.40	1	0	0	0.60	Front(nv)
Oct 1	a-2	0.15	0.10	0	1	0	0.30	Trough
Oct 1	a-3	0.40	0.30	0	0	1	0.45	Stratocu
Oct 1	a-4	0.25	0.25	1	0	0	0.45	Front(v)
Oct 1	a-5	0.05	0.05	0	1	0	0.38	Trough
Oct 1	a-6	1.05	0.10	1	0	0	0.15	ITCZ
Oct 1	a-7	0.10	0.10	1	0	0	0.15	Trop Cyc
Oct 4	b-1	0.55	0.25	1	0	0	0.55	Front(nv)
Oct 4	b-2	0.35	0.20	0	0	1	0.57	Stratocu
Oct 4	b-3	0.10	0.05	0	0	1	0.40	Fog
Oct 4	b-4	0.40	0.15	0	0	1	0.35	Stratocu
Oct 4	b-5	0.15	0.15	1	0	1	0.40	Cirrus
Oct 4	b-6	0.03	0.03	1	0	0	0.25	Trop Cyc
Oct 4	b-7	1.10	0.15	1	0	0	0.15	ITCZ
Oct 7	c-2	0.30	0.20	1	0	1	0.45	Front(v)
Oct 7	c-3	0.04	0.03	0	1	0	0.38	Trough
Oct 7	c-4	0.05	0.05	0	0	1	0.45	Fog
Oct 7	c-5	0.08	0.15	1	0	0	0.15	Trop Cyc
Oct 7	c-6	0.25	0.20	1	0	0	0.45	Front(v)
Oct 7	c-7	0.15	0.15	0	0	1	0.40	Cirrus

Table 2 (continued).

Oct 7	c-8	1.00	0.15	1	0	0	0.15	ITCZ
Oct 10	d-1	0.35	0.35	1	0	0	0.60	Cirrus
Oct 10	d-2	0.20	0.20	0	0	1	0.50	Stratocu
Oct 10	d-3	0.25	0.20	0	1	0	0.45	Trough
Oct 10	d-4	0.15	0.10	0	0	1	0.25	Stratocu
Oct 10	d-5	0.15	0.10	1	0	0	0.15	Trop Cyc
Oct 10	d-6	0.65	0.15	1	0	0	0.15	ITCZ
Oct 10	d-7	0.30	0.15	1	0	0	0.15	ITCZ
Oct 13	e-1	0.40	0.35	1	0	1	0.60	Front(nv)
Oct 13	e-2	0.30	0.25	1	0	1	0.50	Front(nv)
Oct 13	e-3	0.15	0.20	0	0	1	0.50	Stratocu
Oct 13	e-4	0.20	0.15	0	0	1	0.28	Stratocu
Oct 13	e-5	0.35	0.25	0	1	0	0.45	Trough
Oct 13	e-6	0.10	0.10	1	0	0	0.23	Trop Cyc
Oct 13	e-7	0.55	0.10	1	0	0	0.15	ITCZ
Oct 13	e-8	0.50	0.15	1	0	0	0.10	ITCZ
Oct 13	e-9	0.10	0.08	1	0	0	0.18	Trop Cyc
Oct 16	f-1	0.30	0.25	0	0	1	0.60	Stratocu
Oct 16	f-2	0.45	0.30	1	0	1	0.60	Front(nv)
Oct 16	f-3	0.25	0.15	0	1	0	0.50	Trough
Oct 16	f-4	0.10	0.10	0	1	0	0.30	Trough
Oct 16	f-5	0.10	0.15	1	0	0	0.25	Trop Cyc
Oct 16	f-6	0.30	0.25	0	0	1	0.38	Stratocu
Oct 16	f-7	0.08	0.10	1	0	0	0.23	Trop Cyc
Oct 16	f-8	0.35	0.10	1	0	0	0.15	ITCZ
Oct 16	f-9	0.55	0.10	1	0	0	0.13	ITCZ
Oct 19	g-1	0.55	0.25	1	0	0	0.55	Front(nv)
Oct 19	g-2	0.30	0.15	0	0	1	0.60	Stratocu
Oct 19	g-3	0.40	0.35	1	0	0	0.60	Front(nv)
Oct 19	g-4	0.40	0.25	0	0	1	0.38	Stratocu
Oct 19	g-5	0.75	0.10	1	0	0	0.15	ITCZ
Oct 22	h-1	0.10	0.10	0	1	0	0.50	Trough
Oct 22	h-4	0.45	0.20	0	0	1	0.35	Stratocu
Oct 22	h-5	0.10	0.13	0	0	1	0.40	Fog
Oct 22	h-6	1.05	0.10	1	0	0	0.15	ITCZ
Oct 25	i-1	0.30	0.30	0	0	1	0.60	Stratocu
Oct 25	i-2	0.50	0.30	1	0	1	0.55	Front(nv)
Oct 25	i-3	0.15	0.10	0	0	1	0.35	Stratocu
Oct 25	i-4	0.40	0.15	0	0	1	0.25	Stratocu
Oct 25	i-6	0.45	0.10	1	0	0	0.10	ITCZ
Oct 25	i-7	0.50	0.10	1	0	0	0.15	ITCZ
Oct 28	j-1	0.25	0.20	0	0	1	0.55	Stratocu
Oct 28	j-2	0.45	0.25	1	0	1	0.50	Front(nv)
Oct 28	j-3	0.40	0.15	0	0	1	0.30	Stratocu
Oct 28	j-4	0.40	0.30	1	0	0	0.40	Front(nv)
Oct 28	j-5	0.45	0.10	1	0	0	0.15	ITCZ
Oct 28	j-6	0.65	0.10	1	0	0	0.10	ITCZ
Oct 31	k-1	0.50	0.30	1	0	0	0.60	Front(nv)
Oct 31	k-2	0.25	0.30	1	0	0	0.55	Front(nv)
Oct 31	k-3	0.85	0.10	1	0	0	0.15	ITCZ
Nov 3	l-1	0.50	0.20	1	0	0	0.45	Front(nv)
Nov 3	l-2	0.35	0.25	0	0	1	0.55	Stratocu

Table 2 (continued).

Nov	3	l-3	0.40	0.30	1	0	0	0.55	Front(nv)
Nov	3	l-4	0.15	0.10	0	0	1	0.38	Fog
Nov	3	l-5	0.40	0.15	0	0	1	0.28	Stratocu
Nov	3	l-6	0.90	0.10	1	0	0	0.15	ITCZ
Nov	6	m-1	0.60	0.25	1	0	1	0.50	Front(nv)
Nov	6	m-2	0.20	0.20	0	0	1	0.55	Stratocu
Nov	6	m-3	0.10	0.10	1	0	0	0.55	Trough
Nov	6	m-4	0.20	0.10	0	0	1	0.30	Stratocu
Nov	6	m-5	0.95	0.10	1	0	0	0.15	ITCZ
Nov	6	m-6	0.25	0.10	0	1	0	0.20	Cirrus
Nov	9	n-1	0.25	0.25	0	0	1	0.55	Stratocu
Nov	9	n-2	0.20	0.60	1	0	1	0.45	Front(nv)
Nov	9	n-3	0.15	0.15	0	0	1	0.35	Fog
Nov	9	n-4	0.95	0.10	1	0	0	0.15	ITCZ
Nov	12	o-1	0.45	0.25	1	0	0	0.50	Front(nv)
Nov	12	o-2	0.25	0.10	1	0	0	0.45	Front(v)
Nov	12	o-3	0.35	0.20	1	0	1	0.40	Front(nv)
Nov	12	o-4	0.10	0.10	0	0	1	0.30	Stratocu
Nov	12	o-5	1.00	0.10	1	0	0	0.15	ITCZ
Nov	15	p-1	0.55	0.30	1	0	0	0.55	Front(v)
Nov	15	p-2	0.20	0.10	0	0	1	0.55	Stratocu
Nov	15	p-3	0.15	0.20	0	0	1	0.55	Stratocu
Nov	15	p-4	0.40	0.30	1	0	1	0.55	Front(nv)
Nov	15	p-5	0.15	0.15	0	0	1	0.30	Stratocu
Nov	15	p-6	1.05	0.20	1	0	0	0.15	ITCZ
Nov	18	q-1	0.50	0.40	0	0	1	0.60	Stratocu
Nov	18	q-2	0.50	0.40	1	0	0	0.55	Front(v)
Nov	18	q-3	0.10	0.10	0	0	1	0.40	Fog
Nov	18	q-4	0.40	0.15	0	0	1	0.30	Stratocu
Nov	18	q-5	0.90	0.15	1	0	0	0.15	ITCZ
Nov	21	r-1	0.55	0.40	1	0	0	0.65	Front(v)
Nov	21	r-2	0.30	0.30	0	0	1	0.50	Stratocu
Nov	21	r-3	0.30	0.25	1	0	1	0.45	Front(nv)
Nov	21	r-4	0.25	0.20	0	0	1	0.30	Stratocu
Nov	21	r-5	0.95	0.10	1	0	0	0.15	ITCZ
Nov	24	s-1	0.45	0.35	1	0	1	0.55	Front(nv)
Nov	24	s-2	0.15	0.20	0	0	1	0.50	Stratocu
Nov	24	s-3	0.40	0.25	1	0	1	0.45	Front(nv)
Nov	24	s-4	0.90	0.20	1	0	0	0.20	ITCZ
Nov	27	t-1	0.25	0.15	1	0	0	0.45	Front(v)
Nov	27	t-2	0.10	0.15	0	0	1	0.45	Stratocu
Nov	27	t-3	0.20	0.35	1	0	0	0.55	Front(nv)
Nov	27	t-4	0.20	0.25	0	0	1	0.55	Fog
Nov	27	t-5	0.20	0.15	0	0	1	0.35	Stratocu
Nov	27	t-6	0.20	0.35	1	0	1	0.40	Front(nv)
Nov	27	t-7	1.05	0.15	1	0	0	0.15	ITCZ
Nov	30	u-1	0.25	0.15	0	0	1	0.50	Stratocu
Nov	30	u-2	0.55	0.40	1	0	1	0.60	Front(v)
Nov	30	u-3	0.15	0.10	0	0	1	0.30	Stratocu
Nov	30	u-4	0.35	0.20	1	0	1	0.45	Front(nv)
Nov	30	u-5	0.15	0.05	0	0	1	0.30	Stratocu
Nov	30	u-6	0.85	0.10	1	0	0	0.15	ITCZ

Table 2 (continued).

Dec	3	v-1	0.25	0.15	1	0	0	0.40	Front(nv)
Dec	3	v-2	0.45	0.30	1	0	1	0.50	Cirrus
Dec	3	v-3	0.20	0.25	0	0	1	0.55	Stratocu
Dec	3	v-4	0.25	0.25	1	0	1	0.50	Front(nv)
Dec	3	v-5	0.30	0.15	0	0	1	0.30	Stratocu
Dec	3	v-6	0.70	0.15	1	0	0	0.15	ITCZ
Dec	6	w-1	0.20	0.15	1	0	1	0.45	Front(v)
Dec	6	w-2	0.20	0.10	0	0	1	0.50	Stratocu
Dec	6	w-3	0.40	0.25	1	0	0	0.55	Front(nv)
Dec	6	w-4	0.20	0.20	0	0	1	0.35	Stratocu
Dec	6	w-5	0.25	0.15	0	1	0	0.30	Cirrus
Dec	6	w-6	0.20	0.15	0	1	0	0.25	Cirrus
Dec	6	w-7	0.55	0.15	1	0	0	0.15	ITCZ
Dec	9	x-1	0.10	0.15	0	0	1	0.50	Stratocu
Dec	9	x-2	0.45	0.30	1	0	1	0.55	Front(nv)
Dec	9	x-3	0.25	0.20	0	0	1	0.30	Stratocu
Dec	9	x-4	0.55	0.15	1	0	0	0.13	ITCZ
Dec	12	y-1	0.20	0.25	0	0	1	0.45	Stratocu
Dec	12	y-2	0.60	0.30	1	0	1	0.45	Front(nv)
Dec	12	y-3	0.50	0.25	0	0	1	0.35	Stratocu
Dec	12	y-4	0.95	0.15	1	0	0	0.15	ITCZ
Dec	15	z-1	0.10	0.15	1	0	0	0.35	Front(nv)
Dec	15	z-2	0.15	0.25	1	0	0	0.40	Front(nv)
Dec	15	z-3	0.10	0.10	0	0	1	0.45	Fog
Dec	15	z-4	0.20	0.35	1	0	1	0.50	Front(nv)
Dec	15	z-5	0.30	0.25	0	0	1	0.40	Stratocu
Dec	15	z-6	0.95	0.15	1	0	0	0.15	ITCZ
Dec	18	!-1	0.20	0.20	1	0	1	0.40	Front(v)
Dec	18	!-2	0.20	0.25	1	0	0	0.35	Trough
Dec	18	!-3	0.30	0.15	0	0	1	0.35	Stratocu
Dec	18	!-4	0.10	0.10	0	0	1	0.45	Fog
Dec	18	!-5	1.05	0.15	1	0	0	0.15	ITCZ
Dec	21	@-1	0.10	0.20	0	0	1	0.45	Stratocu
Dec	21	@-2	0.20	0.25	1	0	1	0.40	Front(nv)
Dec	21	@-3	0.20	0.15	0	0	1	0.35	Stratocu
Dec	21	@-4	0.95	0.15	1	0	0	0.20	ITCZ
Dec	24	#-1	0.25	0.30	1	0	1	0.40	Front(nv)
Dec	24	#-2	0.20	0.10	0	0	1	0.50	Stratocu
Dec	24	#-3	0.10	0.10	0	1	0	0.30	Trough
Dec	24	#-4	0.15	0.10	0	0	1	0.25	Stratocu
Dec	24	#-5	0.60	0.15	1	0	0	0.15	ITCZ
Dec	27	$-1	0.15	0.15	0	1	0	0.45	Trough
Dec	27	$-2	0.50	0.30	1	0	1	0.50	Front(nv)
Dec	27	$-3	0.35	0.10	0	0	1	0.30	Stratocu
Dec	27	$-4	0.90	0.18	1	0	0	0.18	ITCZ
Dec	30	%-1	0.20	0.20	0	1	0	0.45	Cirrus
Dec	30	%-2	0.25	0.20	0	0	1	0.30	Stratocu
Dec	30	%-3	1.05	0.15	1	0	0	0.15	ITCZ

3. Stepwise Discriminant Analysis

The training of a neural net can require a large number of iterations of the back-propagation procedure. The larger the network, the more computations that must be performed in each iteration. Therefore it is advantageous to keep the network as small as possible. Any inputs that do not actually contribute to the classification process (e.g., have weights close to zero) still require computational effort to derive the network.

To avoid the inclusion of such noncontributing inputs it is useful to perform a preliminary analysis of the data set so that such inputs can be screened from the data set. The statistical method used here is the stepwise discriminant analysis program in the Biomedical Computer Programs P-Series (Dixon and Brown, 1979). In discriminant analysis, cases are divided into groups and statistical analysis is used to find classification functions (linear combinations of the variables) that best characterize the differences between the groups. Variables are entered into the functions one at a time, beginning with the one that contributes most toward differentiating the groups and ending when the group separation fails to improve noticeably. The contribution of each variable is measured by a ratio, called "F-to-Enter," of the sum of the squared errors before and after entry into the equations.

A stepwise discriminant analysis was performed on the cases in Table 2. The first variable entered into the equations was "Mult," the presence of multiple clouds (Table 3). As can be seen in Table 3, the "High" input value has no contribution toward discriminating these groups. Therefore the "High" value

Table 3. Order of entry of variables into stepwise discriminant analysis of cases in Table 2.

Entry Number	Variable	F-to-Enter
1	Mult	270.51
2	North	55.27
3	Del-y	29.95
4	Low	25.85
5	Del-x	2.84
Not entered	High	0.00

is removed from the set of inputs in the neural network derivation described in the next section. It should be emphasized, however, that this result applies only to the data set used here. The "High" cloud parameter may prove to be very useful in future attempts to discriminate different classes of features than those examined here.

4. Neural Network Derivation

Before deriving the neural network, the cases were separated into a dependent, training set and an independent, testing set. The number of cases needed for training depends on the network configuration. The minimum required number of training cases (Ntrain) is heuristically determined by the relation

$$Ntrain = (Nin + Nout)*5.0 \qquad (1)$$

where Nin and Nout are the number of inputs and the number of outputs, respectively (S. Sengupta, personal communication). Given that we have five inputs and eight outputs (i.e., feature types in Table 1), the training sample should have (5+8)*5=65 cases. The training set should also include an equal number of cases of each output class. Thus, the 65 case set divided by eight yields 8.125 cases needed of each type. As can be seen in

Table 1, there is an insufficient number of cases of both "Tropical cyclones" and "Cirrus."

If we exclude these cases from consideration, we now have five inputs and six outputs, requiring (5+6)*5=55 training cases. For an equal number of cases of the six types, we now need 9.167 of each. Now the Fog cases must be excluded due to insufficient numbers (Table 1). With five inputs and five outputs, (5+5)*5=50 training cases are needed, and 10 of each type will suffice. There are enough cases in the remaining five categories (Table 1) to train the network. Unfortunately, there are no "Frontal band (with vortex)" cases left for testing, and only two "Trough" cases left.

Since we do indeed want an independent test of all of the feature types classified by the network, these two classes are also excluded. Now we are left with five inputs and three outputs. The training set needs (5+3)*5=40 cases, and there must be at least 13.333 of each type. Thus, we can train the net with 14 cases each of the "Frontal band (no vortex)," "Stratocumulus" and "ITCZ" features which leaves 23, 39 and 22 cases, respectively, for testing (Table 1).

4.1 Three-Output Neural Network

The network configuration used is depicted in Figure 1. The five inputs connect to a hidden layer of seven units. The first hidden layer connects to a second hidden layer containing four units. Finally, the second hidden layer connects to three outputs, each corresponding to one of the three large-scale cloud features. Although not explicitly shown in this figure, bias

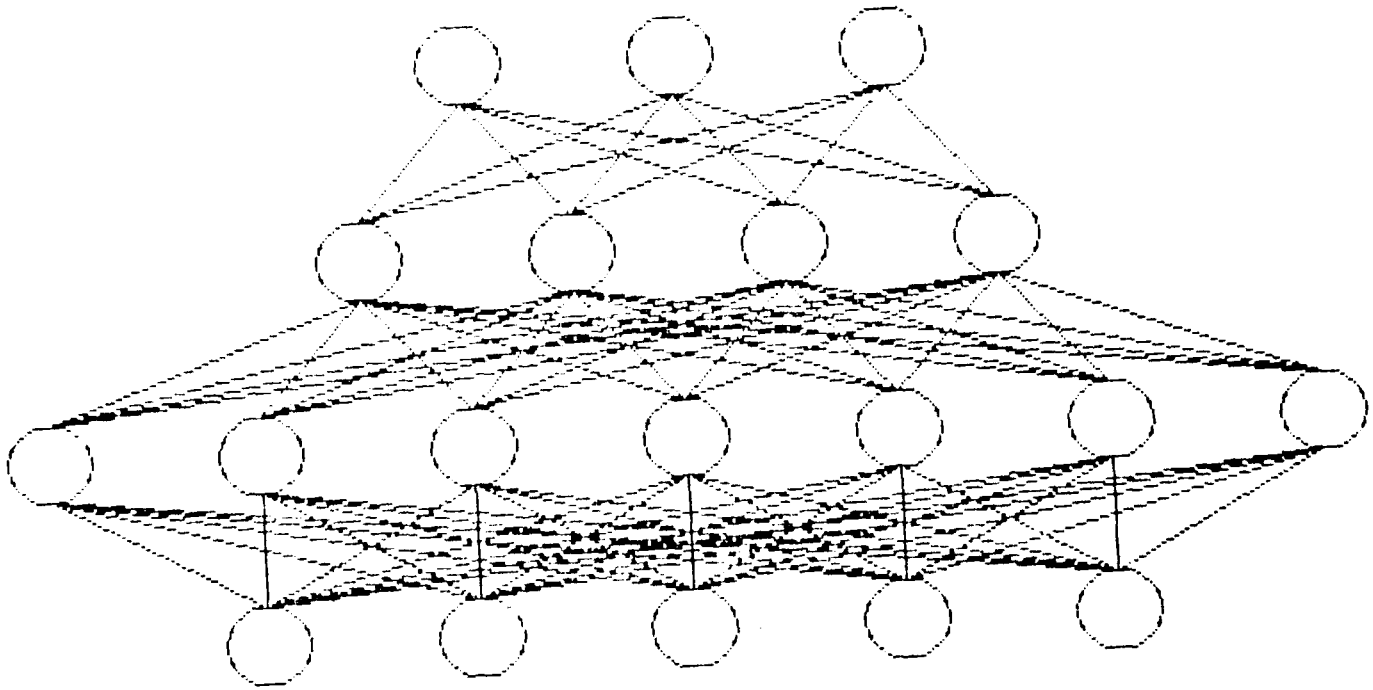


Figure 1. Multi-layer perceptron neural network configuration for the network with five inputs (bottom row) and three outputs top row). Circles represent neural processor nodes and lines denote weight connections between nodes.

terms are also included for all of the hidden-layer units. This network configuration was somewhat arbitrarily chosen; the only consideration used was to increase the number of nodes in the first hidden layer compared to the input layer, and then decrease the number in the second hidden layer to "fan-in" to the output layer. Had the results not been satisfactory, some experimentation in the network configuration would have been tried.

The dependent sample is comprised of the first 14 occurrences in the data set (Table 2) of each of the three features. The network was trained on this 42-case set for 300 cycles before convergence. For this experiment a variable learning rate was used to achieve faster convergence. The initial learning rate was 0.005. This value was sufficient to cause rapid initial decreases in the total sum of the squared errors (tss) between the network outputs and the desired outputs. After about 50 iterations, the tss value had settled down such that it was decreasing by only about 0.001 per cycle. When this occurred, the processing was manually interrupted, the learning rate was increased to 0.01, and training was reinitiated. Whenever the tss value decrease slowed, the learning rate was again gradually increased. In this fashion, convergence was achieved much faster than with a constant learning rate. Network convergence becomes apparent when the tss value begins to oscillate around some low value and no change in the learning rate will cause it to decrease any further. In this experiment the final tss was 0.405.

The performance of the network on the dependent sample cases is indicated in the contingency table presented in Table 4. The

Table 4. Contingency table of dependent sample cases for the network classifying Fronts, Stratocumulus (Strato) and ITCZs. The actual (ACTUAL) classes are presented in the columns while the network-determined (NET) classes are presented in the rows. Tot and Pcnt indicate the totals and percent correct in each line.

		ACTUAL				
		Front	Strato	ITCZ	Tot	Pcnt
N	Front	14	0	0	14	100%
E	Strato	0	14	0	14	100%
T	ITCZ	0	0	14	14	100%
	Tot	14	14	14	42	
	Pcnt	100%	100%	100%		100%

network performs perfectly (100% correct) on these dependent sample cases. When tested on the independent sample cases, the network also performs perfectly (Table 5).

To provide a performance comparison with an alternate technique, discriminant analysis is again used. This time, however, the classification functions are used as classifiers and the results compared with those of the neural net. A discriminant analysis was first run on the dependent sample cases. As can be seen in Table 6, the discriminant analysis classification functions classify 90% of the cases correctly. When applied to the independent sample cases (Table 7) only 86% were classified

Table 5. As in Table 4 except for independent sample cases.

		ACTUAL				
		Front	Strato	ITCZ	Tot	Pcnt
N	Front	23	0	0	23	100%
E	Strato	0	39	0	39	100%
T	ITCZ	0	0	22	22	100%
	Tot	23	39	22	84	
	Pcnt	100%	100%	100%		100%

Table 6. As in Table 4 except for dependent sample classifications using discriminant analysis (DA).

		ACTUAL				
		Front	Strato	ITCZ	Tot	Pcnt
	Front	11	1	0	12	92%
D	Strato	3	13	0	16	81%
A	ITCZ	0	0	14	14	100%
	Tot	14	14	14	42	
	Pcnt	79%	93%	100%		90%

Table 7. As in Table 6 except for independent sample cases.

		ACTUAL				
		Front	Strato	ITCZ	Tot	Pcnt
	Front	12	1	0	13	92%
D	Strato	11	38	0	49	78%
A	ITCZ	0	0	22	22	100%
	Tot	23	39	22	84	
	Pcnt	52%	97%	100%		86%

correctly. Thus, the discriminant analysis technique does not perform as well as the neural network on the same cases (Table 7 vs. Table 5).

4.2 Five-Output Neural Network

Although the above results demonstrate the ability of the neural network approach to the problem of classifying large-scale features, the experiment is limited to only three quite dissimilar types of features. Before the approach can be considered truly applicable to images in an operational environment, it must be demonstrated that it can successfully distinguish between more than three types of large-scale features. For this reason, we return to the analysis at the beginning of Section 4 in which the

requirements of Equation (1) led to paring of the data set. There was a point in the paring process where the potential neural net had five inputs and five outputs, requiring 10 training cases of each type. The five output classes were "Frontal Band (no vortex)," "Frontal Band (with vortex)," "Trough," "Stratocumulus" and "ITCZ." This network was not used at that time because there were not enough "Frontal band (with vortex)" and "Trough" cases left for an independent test of the network. In an effort to provide a net that distinguishes a wider range of classes, it is nevertheless considered useful at this time to derive a neural network using this larger data set. The dependent sample results may in themselves be enlightening. In the author's experience so far with neural networks, their performance seems not to degrade as much when applied to an independent sample as do conventional statistical methods such as regression or discriminant analysis. In addition, the net can still be partially verified using the available independent sample cases. There are 27 "Frontal band (no vortex)," 0 "Frontal band (with vortex)," 2 "Trough," 43 "Stratocumulus" and 26 "ITCZ" cases available for such an independent test.

The network configuration used for this experiment is depicted in Figure 2. As before, there are five inputs leading to seven hidden units. This first hidden layer is connected to a second hidden layer of six units. The output layer has five units, each corresponding to one of the five large-scale feature types.

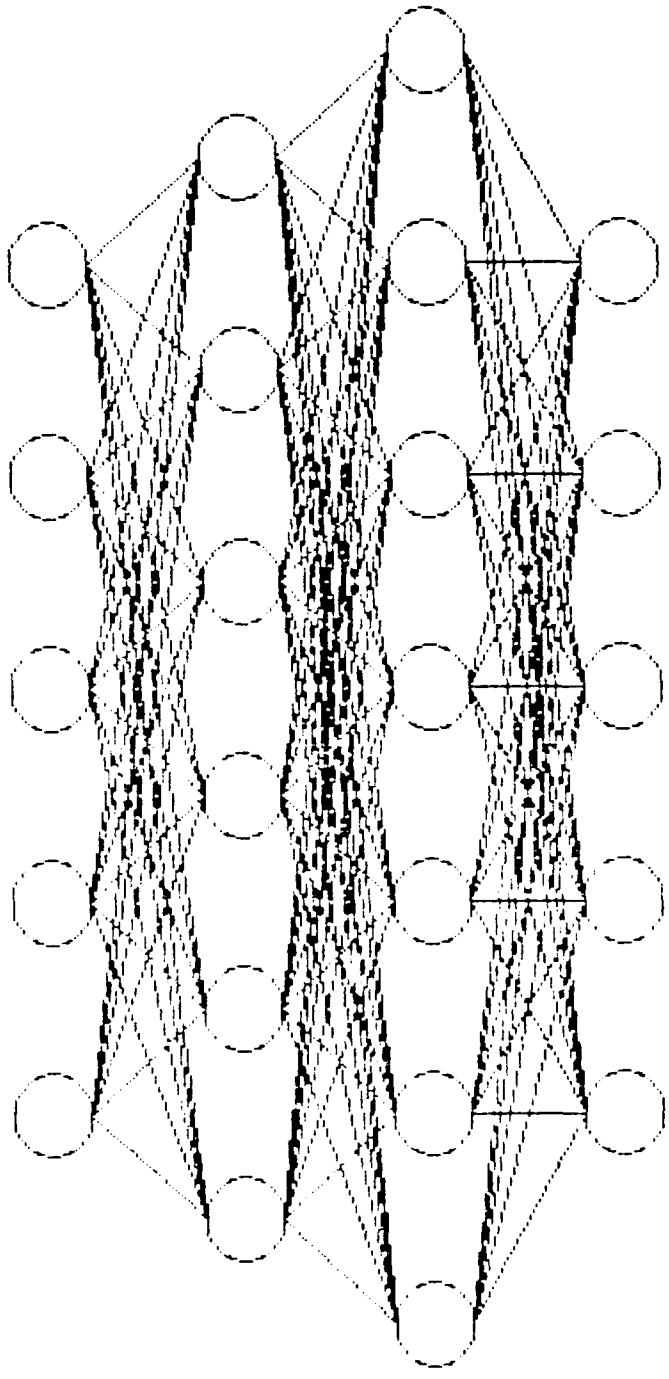


Figure 2. As in Fig. 1 except for neural net with five output nodes.

As in the first network derivation (Section 4.1), the dependent sample is comprised of the first 10 occurrences of each feature type. The network was trained on this 50-case set, again using the technique of a variable learning rate. The network converged to a tss value of 7.66 after 600 iterations.

The network was next verified using the dependent sample cases (Table 8). As in the earlier network results, the "Frontal band (no vortex)," "Stratocumulus" and "ITCZ" classifications are very good (90%, 100% and 100% correct, respectively). The new "Trough" category is also classified with 100% accuracy. The network has difficulty distinguishing "Frontal band (with vortex)" cases (only 60% correct) since they are so similar to the "Frontal band (no vortex)" cases (Table 8), thus lowering the overall accuracy to 90% for the dependent sample. It seems likely that some more sophisticated shape-measurement input would help the net to classify these cases more accurately.

In the independent sample test (Table 9), the "Stratocumulus" and "ITCZ" cases are again classified very accurately (98%

Table 8. As in Table 4 except for the network classifying Frontal bands (without vortices) (Front), Frontal bands (with vortices) (Fr/Vort), Troughs, Stratocumulus (Strato) and ITCZs.

		ACTUAL						
		Front	Fr/Vort	Trough	Strato	ITCZ	Tot	Pcnt
	Front	9	4	0	0	0	13	69%
N	Fr/Vort	1	6	0	0	0	7	86%
E	Trough	0	0	10	0	0	10	100%
T	Strato	0	0	0	10	0	10	100%
	ITCZ	0	0	0	0	10	10	100%
	Tot	10	10	10	10	10	50	
	Pcnt	90%	60%	100%	100%	100%		90%

Table 9. As in Table 8 except for independent sample cases.

		ACTUAL						
		Front	Fr/Vort	Trough	Strato	ITCZ	Tot	Pcnt
	Front	18	0	0	1	0	19	95%
N	Fr/Vort	7	0	0	0	0	7	0%
E	Trough	2	0	2	0	0	4	50%
T	Strato	0	0	0	42	0	42	100%
	ITCZ	0	0	0	0	26	26	100%
	Tot	27	0	2	43	26	98	
	Pcnt	67%	--	100%	98%	100%		90%

and 100% correct, respectively). The two "Trough" cases are also correctly classified. The net has difficulty, though, with the "Frontal band (no vortex)" cases; seven are mistaken as "Frontal band (with vortex)" and two are mistaken as "Trough." Nevertheless, the overall network performance remains at 90% correct. Again, the absence of any "Frontal band (with vortex)" and additional "Trough" cases reduces the significance of this independent sample test.

As in the first experiment, discriminant analysis provides an alternate methodology for comparison. As with the neural net, the "Stratocumulus" and "ITCZ" cases were perfectly classified (Table 10) and the "Frontal band (with vortex)" cases were 60% correct. Two of the "Trough" cases were incorrect, as were four "Frontal band (no vortex)" cases which lowers the accuracy to 80%. Thus, the neural method seems superior to discriminant analysis for the five-category classification as well.

When the discriminant functions were tested on the independent sample, the performance actually increases to 89% (Table 11). This improved independent sample performance is almost certainly

Table 10. As in Table 8 except for dependent sample classifications using discriminant analysis.

		ACTUAL						
		Front	Fr/Vort	Trough	Strato	ITCZ	Tot	Pcnt
	Front	6	4	0	0	0	10	60%
	Fr/Vort	4	6	2	0	0	12	50%
D	Trough	0	0	8	0	0	8	100%
A	Strato	0	0	0	10	0	10	100%
	ITCZ	0	0	0	0	10	10	100%
	Tot	10	10	10	10	10	50	
	Pcnt	60%	60%	80%	100%	100%		80%

Table 11. As in Table 10 except for independent sample cases.

		ACTUAL						
		Front	Fr/Vort	Trough	Strato	ITCZ	Tot	Pcnt
	Front	16	0	0	0	0	16	100%
	Fr/Vort	11	0	0	0	0	11	0%
D	Trough	0	0	2	0	0	2	100%
A	Strato	0	0	0	43	0	43	100%
	ITCZ	0	0	0	0	26	26	100%
	Tot	27	0	2	43	26	98	
	Pcnt	59%	--	100%	100%	100%		89%

due to the independent sample bias toward "Stratocumulus" and "ITCZ" cases. It makes no sense for a statistical method to perform better on independent cases than it does on the developmental set. Of course, it is likely the neural net independent set statistics would also benefit from this bias.

4.3 Eight-Output Neural Network

As a final experiment, we again return to the analysis of the data set with respect to Equation (1). If we are not concerned with an independent sample test, there are nearly enough dependent sample cases to derive a net with eight output classes.

A total of 65 cases (8.125 of each type) is needed. If we assume that eight cases of each type are enough, it can be seen from Table 1 that all except the "Cirrus" category have the required number and the "Cirrus" is only short by one case. With the caveat that the developmental sample may not be sufficient, we can derive a neural net and, hopefully, still learn from the results.

The network configuration used is depicted in Figure 3. Again, there are five inputs leading to seven hidden units. The second hidden layer contains eight units as does the output layer.

As before, the first eight occurrences of each feature type (or all seven "Cirrus" cases) are used to form a 63-cases dependent sample. The network converged with a tss of 15.07 after 600 iterations.

The dependent sample verification is presented in Table 12. The net performs well (83% correct), although it clearly has trouble with the "Frontal band (with vortex)" and "Cirrus" categories (Table 12). Again, a more sophisticated shape measure would likely assist in these classifications.

The independent sample test (Table 13) is again strongly biased toward the inclusion of "Frontal band (no vortex)," "Stratocumulus" and "ITCZ" cases. The network continues to misclassify "Frontal band (with vortex)" cases. Notice also that seven of the "Stratocumulus" cases are misclassified as "Fog" (Table 13). It is clear that some new type of input is needed to separate "Fog" from "Stratocumulus" since both phenomena are of

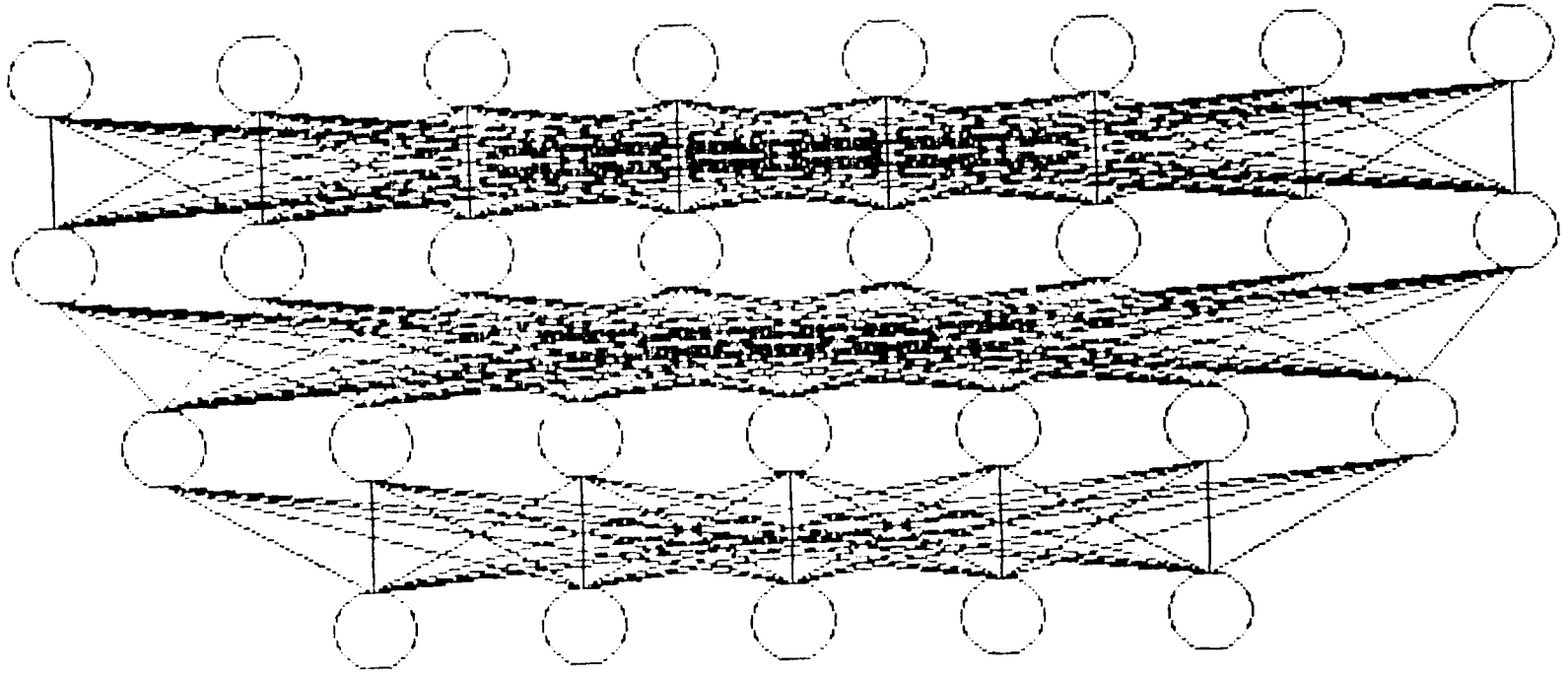


Figure 3. As in Fig. 1 except for neural net with eight output nodes.

Table 12. As in Table 4 except for the network classifying Frontal bands (without vortices) (Frnt), Frontal bands (with vortices) (Fr/Vor), Troughs (Trf), Stratocumulus (Strt), Fog, Tropical Cyclones (TrCy), Cirrus (Cirr) and ITCZs.

		ACTUAL								
		Frnt	Fr/Vor	Trf	Strt	Fog	TrCy	Cirr	ITCZ	Tot	Pcnt
	Frnt	8	5	0	0	0	0	2	0	15	53%
	Fr/Vor	0	3	0	0	0	0	0	0	3	100%
N	Trf	0	0	8	0	0	0	1	0	9	89%
E	Strt	0	0	0	7	2	0	0	0	9	78%
T	Fog	0	0	0	1	6	0	0	0	7	86%
	TrCy	0	0	0	0	0	8	0	0	8	100%
	Cirr	0	0	0	0	0	0	4	0	4	100%
	ITCZ	0	0	0	0	0	0	0	8	8	100%
	Tot	8	8	8	8	8	8	7	8	63	
	Pcnt	100%	38%	100%	88%	75%	100%	57%	100%		83%

Table 13. As in Table 12 except for independent sample cases.

		ACTUAL								
		Frnt	Fr/Vor	Trf	Strt	Fog	TrCy	Cirr	ITCZ	Tot	Pcnt
	Frnt	20	2	0	0	0	0	0	0	22	91%
	Fr/Vor	4	0	0	0	0	0	0	0	4	100%
N	Trf	0	0	3	0	0	0	0	0	3	100%
E	Strt	1	0	0	38	0	0	0	0	39	97%
T	Fog	0	0	0	[illegible]	1	0	0	0	8	13%
	TrCy	2	0	1	0	0	0	0	0	3	0%
	Cirr	2	0	0	0	0	0	0	0	2	0%
	ITCZ	0	0	0	0	0	0	0	28	28	100%
	Tot	29	2	4	45	1	0	0	28	109	
	Pcnt	69%	0%	75%	84%	100%	--	--	100%		83%

similar cloud type and dimensions. The overall percent correct stays the same, however (83%).

As before, discriminant analysis is used for comparison. The dependent sample results (Table 14) show a decrease in overall performance (73% vs. 83%) compared to the neural net, even though the "Frontal band (with vortex)" cases are actually predicted with more skill (Table 12 vs. Table 14).

Table 14. As in Table 12 except for dependent sample classifications using discriminant analysis.

		ACTUAL								
		Frnt	Fr/Vor	Trf	Strt	Fog	TrCy	Cirr	ITCZ	Tot	Pcnt
	Frnt	4	2	0	0	0	0	2	0	8	50%
	Fr/Vor	4	6	0	0	0	0	1	0	11	55%
	Trf	0	0	8	0	0	0	2	0	10	80%
D	Strt	0	0	0	5	1	0	0	0	6	83%
A	Fog	0	0	0	3	7	0	0	0	10	70%
	TrCy	0	0	0	0	0	8	0	2	10	80%
	Cirr	0	0	0	0	0	0	2	0	2	100%
	ITCZ	0	0	0	0	0	0	0	6	6	100%
	Tot	8	8	8	8	8	8	7	8	63	
	Pcnt	50%	75%	100%	63%	88%	100%	29%	75%		73%

In the independent sample test (Table 15), the overall performance decreases only slightly to 72%. The apparent skill in handling the "Frontal band (with vortex)" cases is not apparent in the independent sample test, however. In addition, the discriminant functions have even more difficulty distinguishing "Fog" and "Stratocumulus" than did the neural net (Table 15 vs. Table 13).

Table 15. As in Table 14 except for independent sample cases.

		ACTUAL								
		Frnt	Fr/Vor	Trf	Strt	Fog	TrCy	Cirr	ITCZ	Tot	Pcnt
	Frnt	18	2	0	0	0	0	0	0	20	90%
	Fr/Vor	9	0	1	0	0	0	0	0	10	0%
	Trf	0	0	2	0	0	0	0	0	2	100%
D	Strt	0	0	0	29	0	0	0	0	29	100%
A	Fog	0	0	0	16	1	0	0	0	17	6%
	TrCy	2	0	1	0	0	0	0	0	3	0%
	Cirr	0	0	0	0	0	0	0	0	0	--
	ITCZ	0	0	0	0	0	0	0	28	28	100%
	Tot	29	2	4	45	1	0	0	28	109	
	Pcnt	62%	0%	50%	64%	100%	--	--	100%		72%

4.4 Discussion

The common thread to these results is that the neural network technique is better able to discriminate the feature classes than is discriminant analysis. The reason for this difference is the way each method parses the decision space. In Peak's (1990) Figure 5 and the related discussion, it is shown how different neural net configurations can separate a problem space into various geometric regions. The most complex regions result from the use of nonlinear neural nodes in multiple layers. The discriminant analysis procedure, however, is linear in its combination of input contributions. Thus, the most complex decision regions that can result are convex ones, which are comparable to those defined by a two-layer neural net (Peak, 1990, Fig. 5).

The additional power of a second hidden layer allows neural nets to define concave or even embedded decision regions. Thus, neural nets are inherently superior to discriminant analysis for problems with complex problem spaces.

The set of inputs used in these experiments is probably insufficient for distinguishing such similar features as frontal bands with vs. without vortices, and stratocumulus vs. fog. It would be desirable to have a more complex shape measure than simple zonal/meridional dimensions. For example, a medial-axis transformation might be used to determine the major- and minor-axis lengths of the feature. In addition, actual cloud types would be very useful in place of the simple cloud heights used. Even some measure of the cloudiness density might be used, an indicator which might enable separation of stratocumulus from

fog. The problem with adding more inputs is that the number of training cases required increases by five for each new input. Since the data set was just barely large enough for the experiments presented here, it was not feasible to begin adding new input data types in these experiments.

5. Image Segmentation Considerations

The ultimate goal of this work is to provide an automated image analysis. The data used in the above experiments was acquired only after significant effort by a human interpretation expert in two areas: dividing the image into meaningful, large-scale features and then identifying the cloud types contained in each feature. Automated approaches for cloud-typing are presently under development at NOARL-W. However, the image segmentation problem remains to be addressed. In this section a preliminary experiment in image segmentation will be presented as a possible approach for future research efforts.

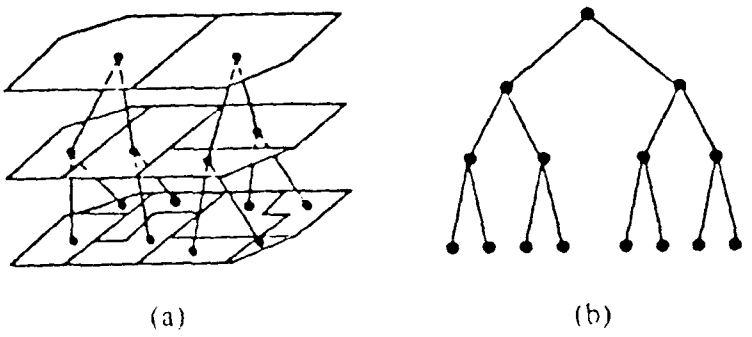
There are two approaches to the segmentation problem. In the first approach, the image is analyzed to find strong gray-scale gradients that correspond to object edges. Once all of the edges are found, the image is separated into regions with common boundaries. The main difficulty with this approach is that edge detection operators not only respond to gradients that actually define region boundaries, but also to gradients that indicate region details or shadows. For images containing regions of nearly the same gray-shade, critical edges may not be detected. In the satellite image problem, adjacent cloud features would be difficult to distinguish in this fashion.

The second segmentation approach involves clustering of regions with similar gray-scales. The analysis begins at the pixel level where some measure of similarity is used to decide which adjacent pixels are most similar. These pixels are combined to form new regions. The process continues as similar adjacent regions are combined until the desired image segmentation is achieved. In this process, it is not the edges that are important, but rather the homogeneity of the interior of each feature.

In the satellite image segmentation problem, cloudy regions have generally lighter gray-scales compared to the darker background ocean or land regions. This characteristic would tend to support the use of the clustering methodology. The unanswered question is what happens when there are adjacent cloud features. Both methodologies may have difficulty in such situations. There may not be well-defined edges when features are adjacent. On the other hand, adjacent cloudy regions may tend to be combined due to similar gray-shades.

The approach presented here is called the Hierarchical Stepwise Optimization (HSWO) algorithm (Beaulieu and Goldberg, 1989). As will be shown, it appears that the region-combining function used by HSWO can accomplish clustering while (hopefully) keeping such adjacent cloudy regions from being combined.

The basis for clustering techniques is the progressive combination of regions, which can be represented by a tree (Fig. 4). In the tree, segments at lower levels are joined to form segments at higher levels. Through a mathematical deriva-



(a) (b)

Figure 4. (a) Segment combination heirarchy during the clustering process (bottom-to-top) and (b) corresponding segment tree (From Beaulieu and Goldberg, 1989).

tion not repeated here, Beaulieu and Goldberg (1989) arrived at a criterion for defining the similarity of adjacent regions. This similarity is defined in terms of the cost of combining adjacent regions:

$$C_{i,j} = \frac{N_i * N_j}{N_i + N_j} (x_i - x_j)^2 \qquad (2)$$

where the subscripts i and j denote adjacent regions i and j, C is the Cost of combining the two regions, N is the number of pixels in a region and x is the mean gray-scale value of the pixels contained in a region. The procedure is to calculate the Cost of combining any two adjacent regions in the image. The two regions that result in the lowest Cost are determined to be the most similar and, therefore, are selected to be combined. Notice that the Cost function is equal to zero when adjacent regions have the same average gray-scale. Thus, the HSWO procedure first combines all of the homogeneous adjacent pixels. As the average gray-scale difference increases, the Cost value rises exponen-

tially. Since the numerator of Equation (2) is the square of the region sizes and the denominator is only their sum, larger regions tend to have higher costs. Thus, the scheme tends to distinguish large-scale regions better than small-scale regions. For the purpose of large-scale feature identification, this property is desirable. The ratio in Equation (2) also ensures that the cost of combining regions of about the same size is higher than the cost of annexing a small region into a large one. It is hoped that this property will cause the HSWO method to distinguish adjacent large-scale features rather than combining them.

The HSWO algorithm is structured to combine the two lowest-cost regions repeatedly until only a single region (the entire image) remains. For a meaningful image segmentation, the merging procedure must be stopped after the noisy, small-scale regions are assimilated but before the meaningful, large-scale regions are combined. Beaulieu and Goldberg (1989) present the example of an image of a checkerboard (Fig. 5). The minimum Cost function value is plotted as a function of the number of segments or iterations (Fig. 6). As the similar regions are combined (following the curve from right to left), the minimum Cost grows gradually. Once the checkerboard squares have all been defined, the system begins to combine them as well. These combinations cause a jump in the minimum Cost function curve (Fig. 6). Thus, the correct stopping point is just before the rapid increase in minimum Cost. Ways to halt the process based on the minimum Cost function increase are a topic for further research and experimen-

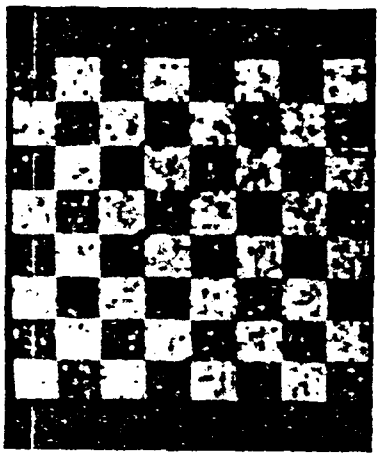


Figure 5. Image of a checkerboard used to demonstrate the HSWO algorithm stopping point (From Beaulieu and Goldberg, 1989).

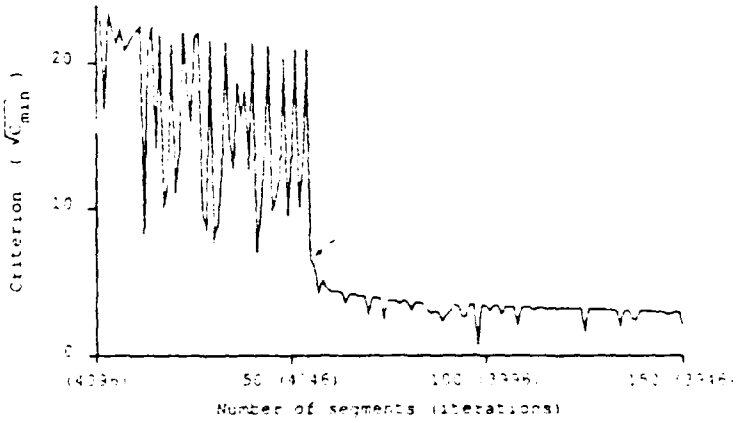


Figure 6. Minimum cost criterion value curve for the checkerboard segmentation problem. Arrow indicated optimum stopping point (From Beaulieu and Goldberg, 1989).

tation, because the shape of the curve depends on the type of image being segmented. It seems reasonable that a satellite image with dark, background regions and bright, cloud features would experience a similar jump that might be detectable as being a good stopping point.

As a demonstration of the HSWO methodology, a satellite image was chosen for segmentation. The image (Fig. 7) is the GOES-W visible image for 2045 UTC on 15 Nov. 1983. The region of interest is the western North Pacific from 105°W to 175°E and from the equator to 55°N. This region contains two frontal bands, two stratocumulus regions and a broad ITCZ. Because the actual gray-scale values are not available, it was decided to use simple percent cloudiness of 5°x5° squares. The visually-estimated cloudiness percentages are presented in Fig. 8.

The HSWO algorithm was programmed in the Prolog language because of its ability to specify the regions dynamically. Initially, each data square is asserted as a Prolog fact containing its cloudiness value and a list of its adjacent neighbors. As the regions are combined, the individual region facts are deleted from the Prolog database and replaced by a new fact representing the combined region, with its new average percent cloudiness and a new, combined list of adjacent regions. In this way, Prolog is a much easier and efficient implementation language than would be one such as C that requires fixed array storage.

At this time, the problem of when to stop the routine is not addressed because the goal of this experiment is to demonstrate the HSWO application to a satellite image. Instead, the evolving segmentation is examined and the process stopped when the image segmentation appears to be at its optimum.

The data values in Fig. 8 were processed by the HSWO program. The resulting regions are depicted in Fig. 9. Here, there

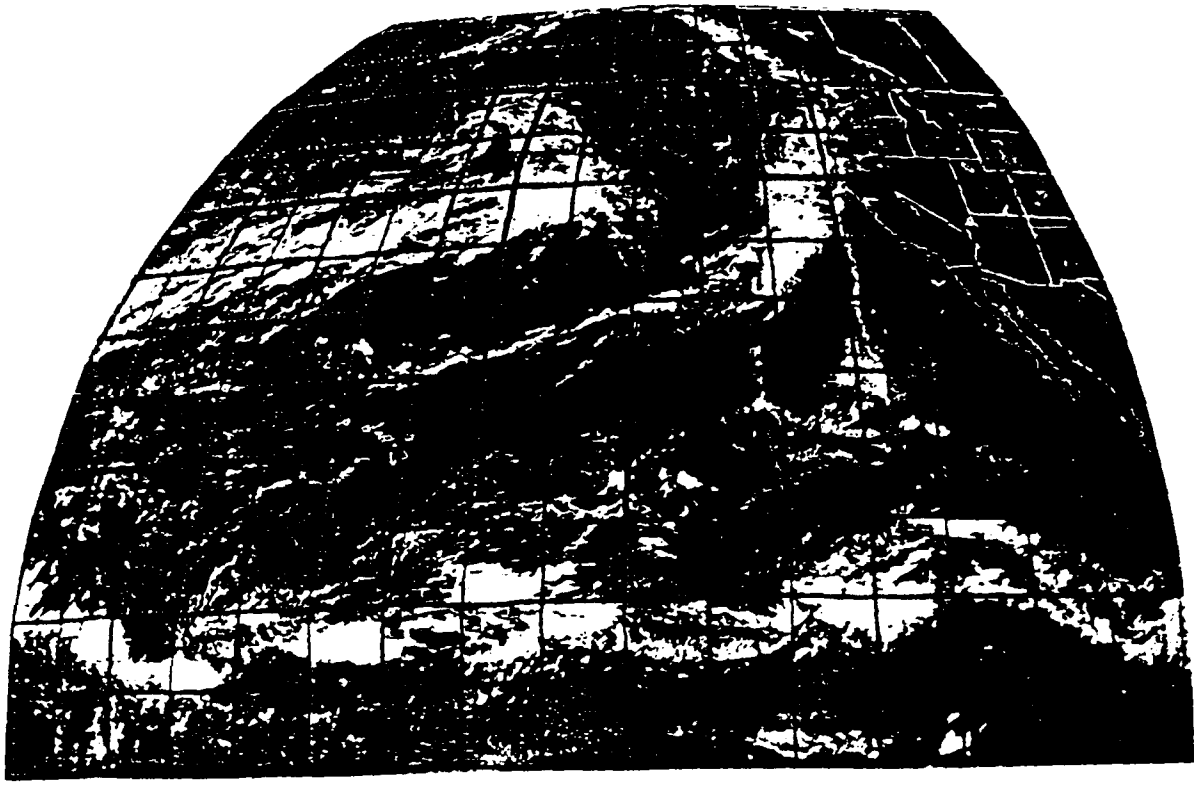


Figure 7. GOES-W visible image used to test the HSWO algorithm. Dark lines define $5^{o}x5^{o}$ squares from which cloudiness percentages are estimated.

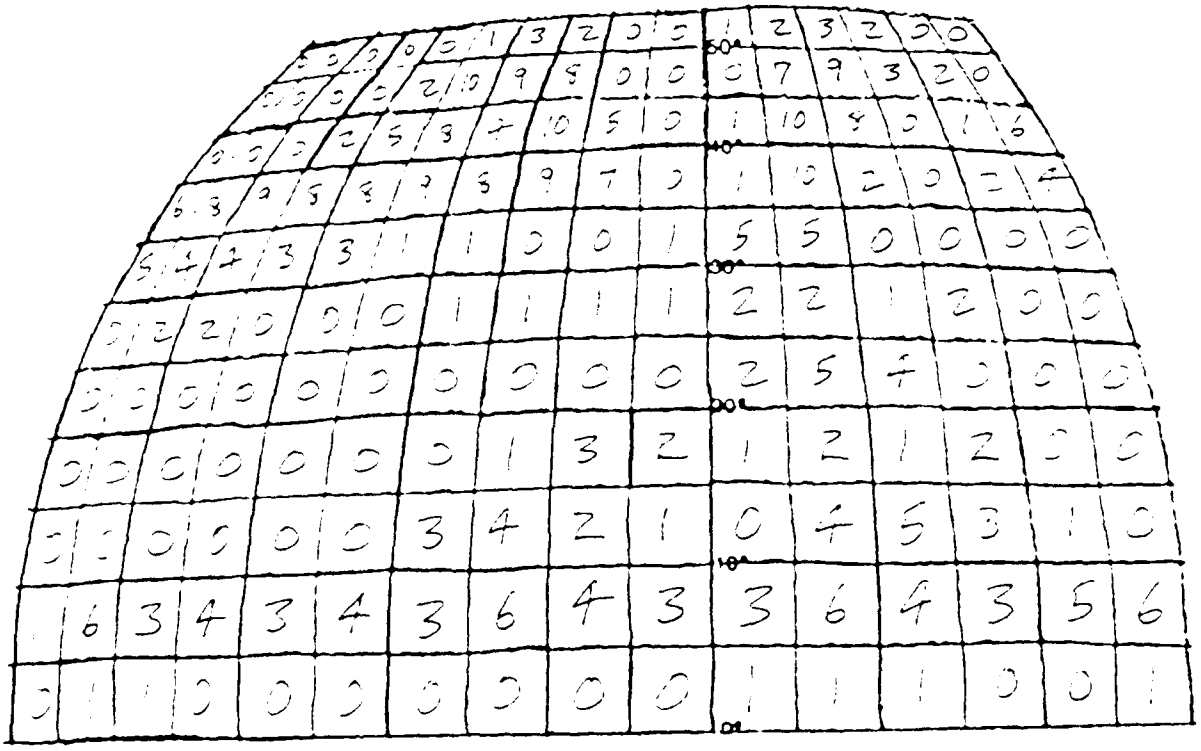


Figure 8. Visually-estimated cloudiness percentages for the image in Fig. 7.

are four distinct cloud features plus a non-cloudy, background region (not numbered in Fig. 9 for clarity). To demonstrate the actual cloud regions distinguished in the image, the four cloud feature segments are overlaid on a noncloudy template (Fig. 10). The long, frontal band in the western Pacific is captured quite well by the algorithm. The main body of the second front in the Pacific Northwest is also captured, but its thin, trailing frontal band was combined into the noncloudy background region rather than into region 2 (Fig. 10). Also, the stratocumulus regions have been lost. It is likely that the use of actual gray-scales and higher resolution would provide a better segmentation of these features. It is interesting that the broad ITCZ is segmented well, but its northward meander from 110°-130°W is not included.

These preliminary results are very encouraging. A digitized gray-scale transform of this image, with 60 pixels-per-inch resolution, was acquired by the author. Unfortunately, there has not yet been enough time to process the data using HSWO. Such a large data set may be too big for the PC-based routine. The availability of Quintus Prolog on the TESS machine would provide the computing power required. Until that Prolog is available, the resolution may have to be reduced by averaging to make the data set more manageable.

6. Conclusions

Three experiments using neural networks to distinguish large-scale cloud features are presented. The data used are taken from GOES-W images from the period October-December 1983.

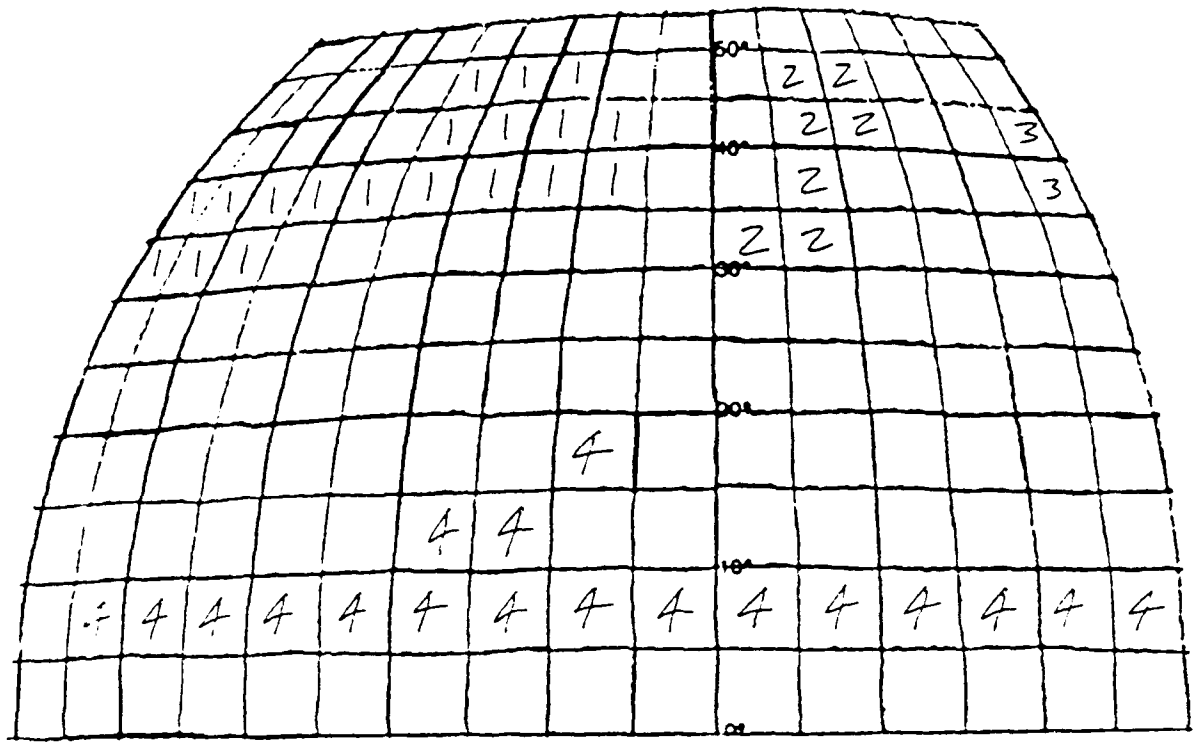


Figure 9. Segmentation regions derived by the HSWO algorithm for the data depicted in Fig. 8. Background region not numbered for clarity.

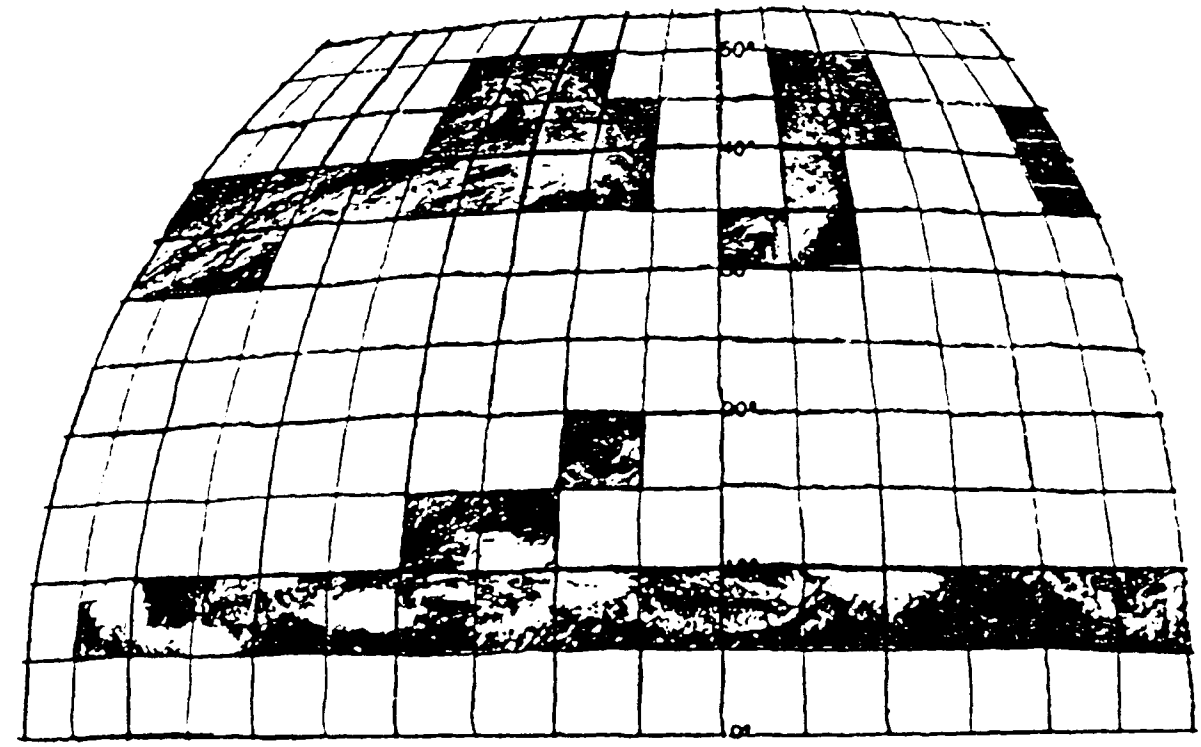


Figure 10. HSWO-derived image segments from Fig. 6 overlaid on the regions in Fig. 9.

Large-scale features and cloud types in the images were categorized by Mr. R. Fett of NOARL-W.

There are eight different cloud features identified on the images. These include "Frontal band (no vortex)," "Frontal band (with vortex)," "Trough," "Stratocumulus," "Fog," "Tropical cyclone," "Cirrus" and "ITCZ." When classified by a neural net, each feature is assigned a different network output node.

The set of five inputs to the network include the zonal and meridional feature dimensions, the presence of multi-level or low clouds, and the north-most latitude of the feature. A potential high cloudiness input was eliminated when discriminant analysis showed that it had no contribution to the discrimination of the feature groups.

When the data set was analyzed to determine the applicable network configurations, it was found that only the "Frontal band (no vortex)," "Stratocumulus" and "ITCZ" features were present in sufficient quantity to provide enough cases for both training and testing a neural network. The network derived to classify these three features is very successful in that all 42 dependent sample and all 84 independent sample cases are correctly classified. This performance compares favorable with the alternate method, discriminant analysis, which could only classify 90% and 86% of the dependent and independent sample cases, respectively.

By foregoing the need for a complete independent sample, the number of classes was expanded to five by adding the "Frontal band (with vortex)" and "Trough" features. The resulting neural network is able to classify 90% of both the 50-case dependent,

and the 98-case independent samples correctly. The independent sample is strongly biased because it contains no "Frontal band (with vortex) and only two "Trough" cases. The discriminant analysis method can only categorize 80% of the dependent sample cases correctly. Discriminant analysis does categorize 89% of the independent sample cases, but this apparent skill is almost certainly anomalous due to the sample bias.

A third experiment was performed in which 65 cases are used to derive a neural net to classify the eight different feature types. The dependent sample results show that the neural net can classify 83% of these cases correctly compared to only 73% for discriminant analysis. Although the independent sample is inadequate for testing this network, the results again indicate superior performance to discriminant analysis (83% correct vs. 72% correct, respectively).

These results indicate that neural networks can classify large-scale cloud features with surprising skill using only very crude input parameters. The eventual inclusion of an automated cloud classification should provide even better input information for future neural net experiments.

The problem of image segmentation is also addressed in this study. A prototype image segmentation routine is developed based on the Hierarchical Stepwise Optimization (HSWO) algorithm of Beaulieu and Goldberg (1989). When tested on a satellite image, the routine seems to be able to segment cloud features while retaining valuable information about their shapes.

Further research in using the HSWO routine is recommended. The goal is to develop this methodology to the point where cloud features can be distinguished. If one of the automated cloud classification routines also becomes available, neural classification experiments similar to those presented here could proceed using automated data exclusively.

REFERENCES

Beaulieu, J.-M. and M. Goldberg, 1989: Hierarchy in picture segmentation - a stepwise optimization approach. IEEE Trans. on Patt. Anal. and Machine Intell., 11, 2, pp 150-163.

Dixon, W.J. and M.B. Brown, eds., 1979: Biomedical Computer Programs P-Series. University of California Press, Berkeley, CA 94720. 770 pp plus appendices.

Peak, J.E., 1991: Neural Network Methodologies and their Potential Application to Cloud Pattern Recognition. NOARL Technical Note 103, Naval Oceanographic and Atmospheric Research Laboratory, Atmospheric Directorate, Monterey, CA 93943-5006. 52 pp.

DISTRIBUTION LIST

SPAWARSYSCOM
ATTN: PMW-141
WASHINGTON, DC 20363-5100

NOARL (10)
ATTN: CODE 125L
JCSSC, MS 39529-5004

NOARL
ATTN: CODE 125P
JCSSC, MS 39529-5004

NOARL
ATTN: CODE 300
JCSSC, MS 39529-5004

OFFICE OF NAVAL RESEARCH
ATTN: CODE 10
800 N. QUINCY ST.
ARLINGTON, VA 22217-5000

WOODS HOLE OCEANOGRAPHIC INST.
ATTN: AI APPLICATIONS
P.O. BOX 32
WOODS HOLE, MA 02543

SCRIPPS INST. OF OCEANOGRAPHY
ATTN: AI APPLICATIONS
BOX 6049
SAN DIEGO, CA 92106

OFFICE OF NAVAL TECHNOLOGY
ATTN: DR. P. SELWYN, CODE 20
800 N. QUINCY ST.
ARLINGTON, VA 22217-5000

OFFICE OF NAVAL TECHNOLOGY
DR. M. BRISCOE, CODE 228
800 N. QUINCY ST.
ARLINGTON, VA 22217-5000

OFFICE OF NAVAL RESEARCH
ATTN: CODE 12
800 N. QUINCY ST.
ARLINGTON, VA 22217-5000

OFFICE OF NAVAL RESEARCH
ATTN: HEAD, OCEAN SCIENCES DIV
CODE 1122
ARLINGTON, VA 22217-5000

U.S. NAVAL ACADEMY
ATTN: LIBRARY REPORTS
ANNAPOLIS, MD 21402

U.S. NAVAL ACADEMY
ATTN: OCEANOGRAPHY DEPT.
ANNAPOLIS, MD 21402

NAVAL POSTGRADUATE SCHOOL
ATTN: CODE MR
MONTEREY, CA 93943-5000

NAVAL POSTGRADUATE SCHOOL
ATTN: 0142
MONTEREY, CA 93943-5002

SPAWARSYSCOM
ATTN: CODE 312
NAT. CTR. #1
WASHINGTON, DC 20363-5100

SPAWARSYSCOM
ATTN: CODE PMW-141
NAT. CTR. #1
WASHINGTON, DC 20363-5100

NAVOCEANSYSCEN
ATTN: J. RICHTER, CODE 54
SAN DIEGO, CA 92152-5000

PACMISTESTCEN
ATTN: GEOPHYSICS OFFICER
PT. MUGU, CA 93042

AFGWC/DAPL
ATTN: TECH. LIBRARY
OFFUTT AFB, NE 68113

AFGL/LY
ATTN: MET. OFFICER
HANSCOM AFB, MA 01731

COMMANDER/DIRECTOR
ASL, WHITE SANDS
ATTN: SLCAS-SE
WSMR, NM 88002-5501

NOAA-NESDIS LIAISON
ATTN: CODE SC2
NASA-JOHNSON SPACE CENTER
HOUSTON, TX 77058

DIRECTOR
NATIONAL EARTH SAT. SERV/SEL
ATTN: FB-4, S321B
SUITLAND, MD 20233

FEDERAL COORD. FOR METEORO.
SERVS. & SUP. RSCH. (OFCM)
ATTN: OPERATIONS SECTION
11426 ROCKVILLE PIKE
ROCKVILLE, MD 20852

DIRECTOR, NOAA
ATTN: TECH. DEVELOPMENT LAB
GRAMAX BLDG.
8060 13TH ST.
SILVER SPRING, MD 20910

NCAR
ATTN: LIBRARY ACQUISITIONS
P.O. BOX 3000
BOULDER, CO 80307

NATIONAL SCIENCE FOUNDATION
ATTN: ATMOS. SCIENCES DIV.
1800 G STREET, NW
WASHINGTON, DC 20550

SCRIPPS INSTITUTION OF
OCEANOGRAPHY, LIBRARY
ATTN: DOCUMENTS/REPORTS SECT.
LA JOLLA, CA 92037

WOODS HOLE OCEANO. INST.
ATTN: DOCUMENT LIBRARY LO-206
WOODS HOLE, MA 02543

COLORADO STATE UNIVERSITY
ATTN: ATMOS. SCIENCES DEPT
ATTN: LIBRARIAN
FT. COLLINS, CO 80523

UNIVERSITY OF WASHINGTON
ATTN: ATMOS. SCIENCES DEPT.
SEATTLE, WA 98195

PENNSYLVANIA STATE UNIV.
ATTN: METEOROLOGY DEPT.
503 DEIKE BLDG.
UNIVERSITY PARK, PA 16802

AMERICAN METEORO. SOCIETY
ATTN: METEOR. & GEOASTRO. ABST
P.O. BOX 1736
WASHINGTON, DC 20013

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. Agency Use Only (Leave blank).	2. Report Date.	3. Report Type and Dates Covered.
	March 1991	Final

4. Title and Subtitle.
Application of Neural Networks to Large-Scale Cloud Pattern Recognition

5. Funding Numbers.
Program Element No. 62435N
Project No. RM35G82
Task No. 4
Accession No. DN651750

6. Author(s).
J.E. Peak

7. Performing Organization Name(s) and Address(es).
Computer Sciences Corporation, Monterey, CA 93943-5006

Naval Oceanographic and Atmospheric Research Laboratory
Atmospheric Directorate, Attn Dr. P. Tag
Monterey, CA 93943-5006

8. Performing Organization Report Number.
NOARL Technical Note 104

9. Sponsoring/Monitoring Agency Name(s) and Address(es).
Office of Naval Research (ONT)
800 N. Quincy St.
Arlington, VA 22217-5000

10. Sponsoring/Monitoring Agency Report Number.
NOARL Technical Note 104

11. Supplementary Notes.

12a. Distribution/Availability Statement.
Approved for public release; distribution is unlimited.

12b. Distribution Code.

13. Abstract (Maximum 200 words).

A Multi-layer Perceptron Neural Network methodology is used to classify eight types of large-scale cloud patterns. The data are taken from GOES-W visible images from Oct. 1-Dec. 31, 1983. Large-scale features are previously identified by a human expert to provide a data set for supervised learning. Discriminant Analysis is used to reduce the set of network inputs and as a comparison classification methodology. In three different tests, the neural network technique classifies the cases with consistently higher accuracy than Discriminant Analysis. The problem of image segmentation is addressed in a preliminary test of the Hierarchical Stepwise Optimization algorithm.

14. Subject Terms.
Neural network
Artificial intelligence
Cloud recognition
Satellite imagery

15. Number of Pages.
45

16. Price Code.

17. Security Classification of Report.	18. Security Classification of This Page.	19. Security Classification of Abstract.	20. Limitation of Abstract.
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	Same as report